DOCUMENT RESUME

ED 040 020 RE 002 791

AUTHOR Farr, Roger
TITLE The Fallacies of Testing.
PUB DATE Mar 70
NOTE 10p.; Paper presented to the Conference on Reading and the National Interest, Bloomington, Ind., Mar. 22-24, 1970

EDRS PRICE EDRS Price MF-$0.25 HC-$0.60
DESCRIPTORS Evaluation, Reading Level, *Reading Tests, *Standardized Tests, Test Construction, *Testing Problems, *Test Interpretation, Test Results

ABSTRACT

Three major points covered by this report are (1) What are the demands for reading assessment and how have the demands increased?; (2) How adequately do present standardized reading tests meet these demands?; and (3) What possible approaches exist for developing assessment procedures which meet these demands? An extreme interest in finding out how well students are reading exists within the educational profession and general public, the author states. Standardized tests are being used extensively to determine the students' reading levels, but almost all of those tests examined showed they are neither able nor designed to meet the demands of the decision situations in which they are often being used. In addition, many situations exist in which the results of these standardized reading tests are being misused and misinterpreted. The author concludes with four basic approaches which he believes may be considered to develop assessment procedures that meet the demands for accurate measurement of reading achievement. (NH)

ED 040020

U.S. DEPARTMENT OF HEALTH, EDUCATION
& WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

THE FALLACIES OF TESTING

PRESENTED BY

DR. ROGER FARR

DIRECTOR, READING CLINIC
AND
CO-EDITOR READING RESEARCH QUARTERLY

TO THE

CONFERENCE

ON

READING AND THE NATIONAL INTEREST

Indiana University
Bloomington, Indiana

March 22-24, 1970

RE 002 791

Before I begin my comments on reading tests, I want to assure you that I am neither attempting to foster a "Ban the Test movement," nor claiming that the assignment of a number to an event is the end goal of any education effort. Rather, I believe that we should consider testing as only one means of making better instructional decisions. This position often brings up the question of what can be measured? Aren't there some things which are intrinsically immeasureable? Abraham Kaplan, in his book The Conduct of Inquiry, answers this question more cogently than I: "For my part, I answer these questions with an unequivocal 'No.' I would say that whether we can measure something depends, not on the thing, but on how we have conceptualized it, on our knowledge of it, and above all on the skill and ingenuity which we can bring to bear on the process of measurement which our inquiry can put to use."[1]

I would like to address my remarks today to three major points:

1. What are the demands for reading assessment and how have the demands increased?
2. How adequately do present standardized reading tests meet these demands?
3. What possible approaches exist for developing assessment procedures which meet these demands?

Demands and reasons for more valid reading assessment

The plea for more valid measurement of reading behaviors is not new. However, the emphasis in Congress on accountability, the attempts by several publishing corporations to sell instructional products on a sliding cost scale based upon reading gains of children, the national Right to Read program, and the targeted research plan of the U.S. Office of Education have all contributed to a growing interest in the assessment of reading

---

[1]Abraham Kaplan. The Conduct of Inquiry, (San Francisco: Chandler Publishing Company, 1964,) P. 176.

behaviors.

In fact, the targeted research program will be developed on the basis of criterion tests of reading. The following quote from a recent announcement of the targeted research program makes the point quite specifically:

"The U.S. Office of Education intends to support a five phase program of research and development on reading to reach the following objectives: 100 percent of all persons not in permanent care institutions must pass, by age 10, a criterion-referenced test which is predictive of competent performance on a set of adult reading tasks selected to have a favorable returns to the individual and to society in general."[1]

Dr. James Allen has also alluded to the development of criterion tests in his speeches and comments on the Right to Read program. In a recent issue of Family Weekly,[2] Dr. Allen stated that one in four students nationwide has significant reading deficiencies and that up to half of the students in large city school systems read below expectations for their age levels. He also stated that among unemployed young people between ages 16 and 21, about half are functionally illiterate.

A consideration of statements and efforts like those above has led me to the following conclusions:

1. There exists among both the education profession and the general public an extreme desire and interest in finding out how well students are reading -- in a very functional way.
2. Standardized reading tests (usually standardized tests developed by large publishing companies) are being used extensively to

---

[1]"Research and Development Sources Sought", Commerce Business Daily, February 25, 1970.

[2]"Goal for the '70's: To Improve Your Child's Reading: An exclusive interview with Dr. James E. Allen, Jr." James C. G. Conniff. Family Weekly, March 15, 1970.

determine how well students are reading.

3. There are many, many situations in which the results of standardized reading tests are being misused and misinterpreted.

Before turning to consideration of how well present standardized reading tests meet the needs that I have delineated, I would like to define the difference between a criterion test and a standardized or norm referenced test. I want to make this distinction clear because the testing needs that exist in the field of reading necessitate the use of criterion referenced tests, but norm referenced tests are now being used to fulfill these needs.

Criterion referenced tests are very closely related to the old concept of a mastery test: the purpose of such a test is to measure achievement of a very specific behavior and often to make a very specific decision. For example, has Bill mastered the skills necessary to drive a car? Is Sam able to swim a mile? Or has Jerry mastered the essential beginning reading skills necessary to go on to the next phase of instruction? In each of these situations, the criterion is quite definite and the student is assessed to determine whether he can complete the task.

A standardized norm referenced test is also concerned with assessing behaviors and making decisions, but the decisions are of a comparative nature. For example, how good a driver is Bill compared with Sam? Is Sam an adequate swimmer for his age and size? Or how good is Jerry's reading skill development compared to other students at his grade level? Another way to consider the basic difference between the two types of tests is to consider the anchor point for each test. A norm referenced test is usually anchored in the middle of the ability of the group to be tested; the test performances will then tend to spread out so comparisons

can be made. A criterion referenced test, on the other hand, is anchored at one end. The test developer is not interested in the spread of performances but rather in how many students are able to perform well enough to pass the anchor point.

As I said earlier, criterion referenced tests are needed to make the decisions that are being posed by the Right to Read program and the Targeted Research effort of the U.S. Office of Education. I also indicated that the results of standardized reading tests are now being used as the data base for making these decisions.

How adequately do standardized reading tests meet these demands?

An examination of almost all existing standardized reading tests and of the research concerned with these instruments leads to the conclusion that these tests are neither able nor were they designed to meet the demands of the decision situations for which they are often being used.

This should not be taken as a blanket rejection of standardized reading tests. Given a clear understanding of their major purpose i.e., the comparison of groups and individuals; and knowledge of the limitations of the tests, they can be very useful tools. However, they do not satisfy the needs I delineated earlier.

Standardized norm referenced tests do not result in any information about what a student can do. For example, we have no basis for deciding what a raw score of 121 points, a grade score of 4.2, or a percentile of 63 means so far as the actual reading tasks a student with such scores would be able to perform. We can only use these scores for comparing the student to some norm group.

Furthermore, it is quite clear that the development of the subtests on almost all standardized reading tests is based upon vague assumptions

about the reading act and that there is no clearly defined or empirically supported evidence to validate the existing subtests of most standardized reading tests. In fact, there is considerable evidence that these existing subtests are not valid measures of actual reading subskills. The subskills problem is confounded even more by the very obvious lack of agreement as to what the actual subskills of reading are and how each should be measured. In fact, one could list several hundred different reading subskills by examining the titles of the subtests of standardized reading tests. In addition, an examination of the numerous approaches to measuring any one of these hundreds of skills further increases one's doubt as to what can be measured. While this talk is not intended to dwell on the uses of standardized reading tests, let me at least suggest that if you use such tests you can feel quite confident about using the total test scores for comparative purposes, but I strongly suggest that you do not use the subtest scores for either diagnostic or comparative purposes.

The preceding statements suggest that most present standardized reading tests cannot meet the needs of criterion decisions. The tests cannot tell us what a student can be expected to do, what skill development he needs, or whether he is meeting the objective of basic literacy. Let me again emphasize that most standardized reading tests were not developed to meet these needs and the manuals of most standardized reading tests suggest that the tests should not be used for these purposes. That caution, however, does not seem to deter test consumers from misusing and misinterpreting the tests.

If present standardized reading tests do not meet our criteria decision needs, what approaches exist for developing assessment procedures that 'successfully' meet these needs? From my thinking on this topic I have

concluded that there are four basic approaches which may be considered. These are: (1) the development of assessments based on criterion objectives that would reflect the reading demands of an effective citizen; (2) the use of average levels of achievement of some age groups as standards of achievement; (3) the development of norm referenced tests with more specific behavioral objectives built into them; and (4) the development of guidelines that could be used to develop situation specific criterion measures.

Let me expand a bit on each of these approaches. The development of assessments based on criterion objectives is exemplified by such tests as the New York State Minimum Competency Test in Reading. While I am not citing this test as an outstanding example, it presents some of the problems of this approach. The manual of the test states that a score of 26 correct responses out of a total of 40 multiple choice questions based on a series of reading selections is the standard of minimum reading competence for a New York State high school graduate.

Another example of this approach is a reading test which I have been thinking about since the Federal government delivered my personal income tax forms last December. It occurred to me that the personal federal income tax form is perhaps the most widely-read reading material that I could conceptualize. Completion of the forms obviously necessitates a very functional kind of reading, but there is also a general introductory section at the beginning of the forms which seems to necessitate a more general kind of reading. This is the section that discusses why everyone should be a good citizen and pay his taxes; it also indicates the uses to which the government is going to put the tax money and where all the money is going to come from. In addition to the functional reading and general

reading power needed to complete the form, there are graphs and charts to interpret. All in all, I think the federal income tax form might make a very excellent criterion referenced test.

The problems of this first approach are primarily concerned with the arbitrariness of deciding on the criterion tasks. The development of the tasks for the test would always be arbitrary to a large degree, as they were in the two examples I just cited; and many groups would object to the definitions of functional literacy implied by the tasks on the test. In addition, the content of the test could be faulted on the grounds that it is not representative of basic reading ability needed by adults in our society. There would also be concerns raised about the possibility that such a test might limit the reading development of children; this would be a legitimate concern if schools or teachers were satisfied with achieving the basic levels of reading competency represented by the test and did not try to develop each child to his fullest potential.

The advantages of this approach would be that the test could be used for making decisions about the content of reading programs. It could also be used as the focus for a study of reading subskills by encouraging research on discovering the skills necessary to perform adequately on the criterion test. Finally, such tests could be used as the bench mark for determining the number of functional illiterates in the United States.

The second approach would involve the adoption of some average level of reading achievement for an arbitrarily-chosen age-level as the criterion for basic literacy. For example, we could decide that the definition of functional literacy is the reading score the average 15-year-old achieves on a test of general reading achievement, such as the total score on a standard reading test. This approach has the same disadvantages as

traditional standardized reading tests. There are no clear cut objectives built into the test and there is no reference for interpreting the reading performance. The advantage of this approach is that it would be quite easy to develop and would result in an immediate criterion for making decisions about literacy levels.

The third approach would be the development of clearly-stated behavioral objectives within a norm referenced reading measure. This approach would partially combine the first two appraaches I have just described. It would result in a testing instrument which would include criterion references as well as norm references. For example, an eighth grade student might get a raw score of 121 points on the test. This might be interpreted to mean that he has the necessary functional reading skills to complete his personal income tax forms and also that he is reading as well as the average ninth grader in his second semester, or a grade score of 9.7.

The problems of developing such a test would be monumental. I do think we have the measurement knowledge and ingenuity to produce such a test, but I do not believe we know enough about reading behaviors to be very successful. Much research is needed before we can go further. I think the first thing we need to know is what the reading demands of particular economic groups are. A doctoral student here at Indiana is now developing a study in which he intends to examine the reading demand of specified occupations and whether the employees in that occupation have the necessary reading skills for the reading demands of their jobs. He also intends to study the non-occupation reading habits of these people. I think that a series of such studies would provide much needed information for deciding on the criterion objectives for the makeup of literacy tests.

Another avenue of research leading to the development of this third approach is more extensive observations of the reading behaviors of subjects engaged in reading. These need to be conducted for a variety of reading tests under a variety of conditions. Some of the work of psycholinguists, such as Ken Goodman of Wayne State University, offer initial leads along this line.

"If" we are to develop criteria tests, I would like to see them developed so that they met the definition of this third approach and that they be based on more systematic study of reading behaviors. However, my fourth approach suggests that it may not be feasible or logical to attempt to develop national criterion tests in reading. It may be that each situation demands the need for its own criterion test. Varying socio-economic, geographical or community objectives may block the development of national criteria for basic literacy. If this is the case, we need not be immobilized. We could develop guidelines and training programs for the development of situation-specific criterion tests. These guidelines would cover such topics as defining behaviors, identifying goals, developing behavioral objectives, sampling behaviors, and test analysis concepts.

Each of the four procedures which I have briefly explored offers a possibility for meeting the reading assessment needs of the nation. There are limitations and problems inherent in each approach; the approaches that seem to offer the best alternative involve the most extensive effort, but if we are sincerely dedicated to the Right to Read Program, we need to face the assessment problem at the outset. Facing this problem, will not only tell us where we now are, but will also force us to consider where we want to go.

Summary:

I would like to summarize briefly my three major points. First, I think there exists a vital need for valid criterion referenced measures in reading. In order to make valid decisions we need valid assessment data. Second, there are almost no testing instruments available today which can fulfill these needs. Standardized reading tests are presently being misused for these needs and this is leading to some rather unfortunate conclusions and decisions in the field of reading. Third, I think there are several approaches that can lead to the development of the kind of tests we need.

A consideration and study of these approaches and any others that might be added to my list of four should be given immediate attention and top priority at the beginning of the Right to Read effort. If this is not done there will be little means of directing efforts toward what we want to accomplish, and no way of knowing whether we achieve our goals.